Data Analysis

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- The original dichotomous discriminant analysis was developed by Sir Ronald Fisher in 1936.
- It is different from an ANOVA or MANOVA, which is used to predict one (ANOVA) or multiple (MANOVA) **continuous dependent variables** by one or more **independent categorical variables**.
- Discriminant function analysis is useful in determining whether a set of variables is effective in predicting category membership.

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- LDA works when the measurements made on independent variables for each **observation are continuous quantities**. When dealing with categorical independent variables, the equivalent technique is discriminant correspondence analysis.
- LDA is also closely related to principal component analysis (PCA) and exploratory factor analysis (EFA) in that they both look for linear combinations of variables which best explain the data.
- LDA explicitly attempts to model the difference between the classes of data. PCA does not take into account any difference in class, and EFA builds the feature combinations based on differences rather than similarities.

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- It is also known as factorial discriminant analysis (FDA) or canonical discriminant analysis (CDA).
- The analysis perspective is similar to that of principal component analysis (PCA).
- Factorial discriminant analysis aims to determine 2 new predictive variables, named discriminant variables, such that the individual observations to be separated as clear as possible, based on these variables.
- The discriminant variables are, as in PCA, linear combinations of initial variables (from matrix X), and uncorrelated to each other.

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- A natural criterion for determining discriminant variables is to maximize class or group cohesion, based on the intra-class and inter-class variance, i.e. the ratio between the inter-class variance of a variable and the total variance (or the intra-class variance) to be as great as possible.
- *Let's assume that the observation matrix*, X, *is centered.*
- Then the first discriminant variable is determined as follows:

$$z_1 = \begin{bmatrix} z_{11} \\ z_{21} \\ \dots \\ z_{n1} \end{bmatrix} = X \cdot u_1$$

, where u_1 is the first discriminant factor.

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- The u_1 coefficients of the linear combination are:

$$u_1 = \begin{bmatrix} u_{11} \\ u_{21} \\ \dots \\ u_{m1} \end{bmatrix}.$$

- Group centers for variable z_1 are: $\overline{z_1} = \begin{bmatrix} \overline{z}_{11} \\ \overline{z}_{21} \\ \dots \\ \overline{z}_{q1} \end{bmatrix} = G \cdot u_1$,

where G = $\begin{bmatrix} g_{11} & g_{12} & \dots & g_{1m} \\ g_{21} & g_{22} & \dots & g_{2m} \\ \dots & & & \\ g_{q1} & g_{q2} & \dots & g_{qm} \end{bmatrix}$,

g_{kj} representing the mean of predictor variable j for the k group.

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- If X is not a centered matrix, then $z_1 = u_{01} + X \cdot u_1$.
- The discriminant variables can be viewed as discriminant functions (*Figure 1*), also named Fisher functions, whereby a good separation among instances (observations) can be made.
- z_1 is a hyperplane with of equation $u_{01} + X \cdot u_1$, which split the instances (continuous red line).
- u_1 is the perpendicular line on the hyperplane, or the discriminant axis (dashed red line) which goes through the origin, if the data is centered, or through the center of gravity (the black dot), if the data is not centered.

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

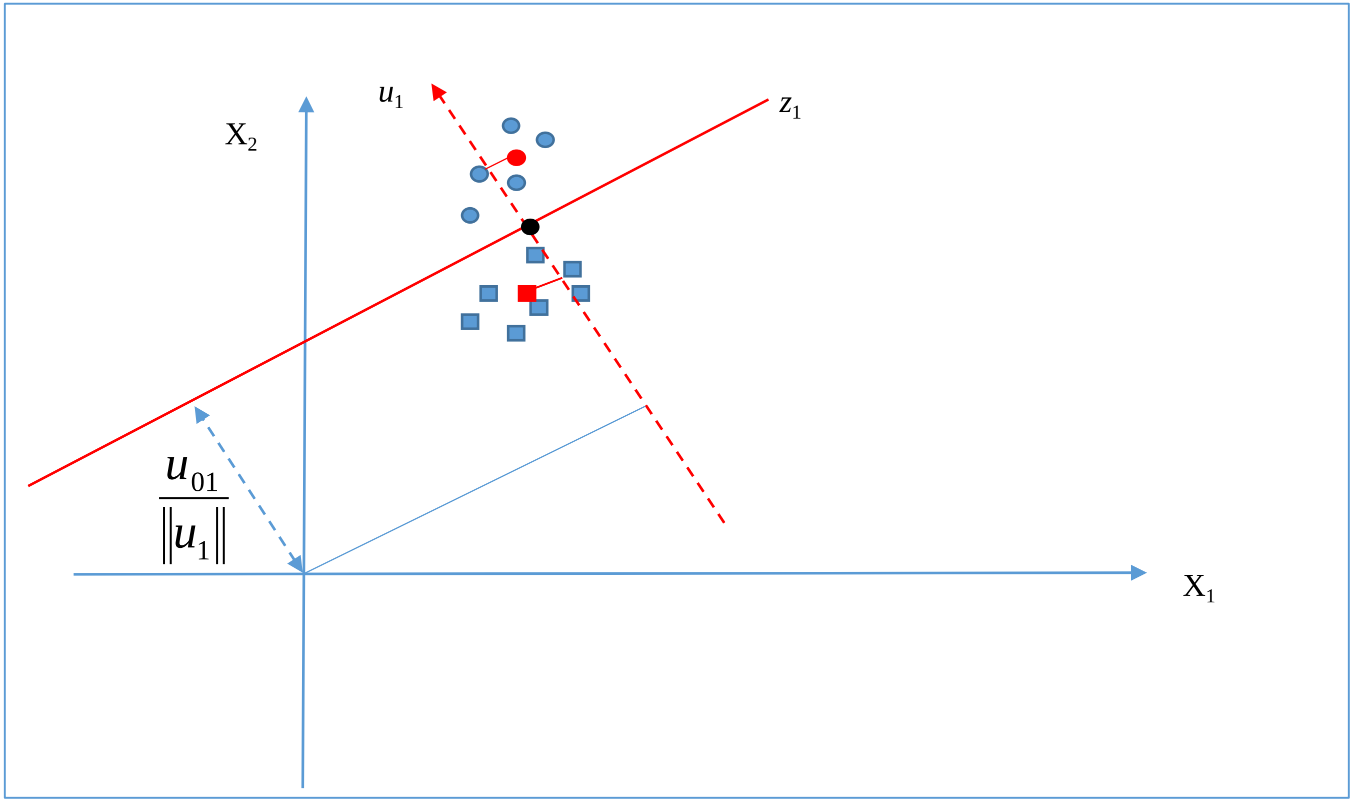


Figure 1*. Fisher discriminant functions. Bi-dimensional case, with two classes.*

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- If the data is centered, the free term u_{01} is 0 (zero).
- u_1 axis is chosen such that to separate the groups as clear (decisive) as possible, i.e. the distances between the group centers to the axis is to be as great as possible (the group centers are the red dots).
- Total variance of variable z_1, in terms of matrices is:

$$VT_1 = \frac{1}{n}\left(z_1\right)^t z_1 = \frac{1}{n}\left(X \cdot u_1\right)^t X \cdot u_1 = \frac{1}{n}\left(u_1\right)^t X^t X \cdot u_1 \ ,$$

where $(u_1)^t$ is a row vector, the transposed of column vector u_1.

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- The inter-class variance of variable z_1 is:

$$VB_1 = \sum_{k=1}^{q} \frac{n_k}{n} (\overline{z}_{k1})^2,$$ where $\overline{z}_{k1}$ is the mean of the variable for class k, and n_k is the no. of observations belonging to class k.

- In terms of matrices, the inter-class variance can be written as:

$$VB_1 = \frac{1}{n}\left(\overline{z}_1\right)^t D_G \cdot \overline{z}_1 = \frac{1}{n}\left(G \cdot u_1\right)^t D_G G \cdot u_1 = \frac{1}{n}\left(u_1\right)^t G^t D_G G \cdot u_1 ,$$

where D_G is diagonal matrix of group frequencies: $D_G = \begin{bmatrix} n_1 & 0 & \ldots & 0 \\ 0 & n_2 & \ldots & 0 \\ \ldots & & & \\ 0 & 0 & \ldots & n_q \end{bmatrix}$.

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- The intra-class variance is given by:

$$VW_1 = \sum_{k=1}^{q} \frac{n_k}{n} \cdot \frac{1}{n_k} \sum_{i \in k} \left(z_{i1} - \overline{z}_{k1} \right)^2 \quad .$$

- Where n_k is the number of observations belonging to class (group) k,
- and the relation between variances is: $VT_1 = VB_1 + VW_1$.

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- Having a given value for the total variance, a variable discriminates better the classes (groups) if the following conditions were better satisfied:
  - the instances belonging to a class have as close as possible values, i.e. the intra-class variance is minimal;
  - the mean of the classes are as far apart from each other as possible, i.e. the inter-class variance is maximal;
  - the discrimination power of a variable is the ratio between inter-class variance and total variance.

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- Maximizing the distance between the center projections on the discriminant axis u_1 is equivalent with maximizing the ratio: $\frac{VB_1}{VT_1}$ for the discriminant variable z_1.
- If we label this ratio with α_1, then we have:

$$\alpha_1 = \frac{VB_1}{VT_1} = \frac{\frac{1}{n}(u_1)^t G^t D_G G \cdot u_1}{\frac{1}{n}(u_1)^t X^t X \cdot u_1}$$

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- Then the optimum problem is:

$$\left\{ \underset{u_1}{Maxim} \frac{\frac{1}{n}(u_1)^t G^t D_G G \cdot u_1}{\frac{1}{n}(u_1)^t X^t X \cdot u_1} \right.$$, where:

u_1 is the discriminant factor,

$B = \frac{1}{n} G^t D_G G$ is the inter-class covariance matrix, and

$T = \frac{1}{n} X^t X$ is the total covariance matrix.

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- The solution u_1, obtained by solving the optimum problem, is the eigenvector of $T^{-1}B$ matrix, corresponding to the greatest eigenvalue.
- This eigenvalue is actually α_1, the discrimination power of variable z_1.
- The next discrimination variables are determined in the same manner, having the following additional conditions of not being correlated at all to the previously determined discriminant variables:

$$\frac{1}{n}\left(z_k\right)^t z_j = 0 \;, \quad j = 1, k\text{-}1.$$

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- The optimum problem becomes:

$$\begin{cases} \underset{u_k}{Maxim} \dfrac{\frac{1}{n}(u_k)^t G^t D_G G \cdot u_k}{\frac{1}{n}(u_k)^t X^t X \cdot u_k} \\ \frac{1}{n}(u_k)^t X^t X \cdot u_j = 0, \quad j = \overline{1, k-1} \end{cases}$$

- The solution of the problem, u_k factor, is the eigenvector of $T^{-1}B$ matrix, corresponding to the eigenvalue of order k (in the descending order of the eigenvalues) α_k.
- α_k is the discrimination power of variablei z_k.

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- The number of the discriminant variables is given by the number of not null eigenvalues of $T^{-1}B$ matrix (see the canonical discriminant analysis below), and this number is:

  r = minim (m, q-1).
- The solutions of the model are:

$$T^{-1}B \cdot u_k = \alpha_k \cdot u_k \,,\ \ k = 1, r$$

$$z_k = X \cdot u_k \,,\ \ k = 1, r$$

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- The discrimination power of discriminant variables can be also computed as ratio between inter-class and intra-class variance.
- Consequently, the optimum criterion to obtain the discriminant variables is maximize the ratio $\frac{VB_k}{VW_k}$.
- Computed in this manner, the discrimination power of the variable is greater, since:

$$\frac{VB_k}{VW_k} \geq \frac{VB_k}{VB_k + VW_k}$$

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- Labeling $\frac{VB_k}{VW_k} = \lambda_k$, the discrimination power of discriminant variable z_k, there is:

$$\lambda_k = \frac{VB_k}{VW_k} = \frac{VB_k}{VT_k - VB_k} = \frac{VB_k}{VT_k \cdot \left(1 - \frac{VB_k}{VT_k}\right)} = \frac{\alpha_k}{1-\alpha_k}$$

- The values λ_k, $k = 1,r$ are not null eigenvalues of $W^{-1}B$ matrix, corresponding to the same eigenvectors, u_k.

Discriminant Analysis

Linear Discriminant Analysis (LDA). Fisher functions.

- The solutions of the modified model are:

$$W^{-1}B \cdot u_k = \lambda_k \cdot u_k,\ \ k = 1, r$$

$$z_k = X \cdot u_k\ , k = 1, r.$$

Discriminant Analysis

Result graphical representations.

- *Observation representations*. A synthetic image of those *n* instances (individuals) distributed in *q* groups can be obtained employing a 2D or a 3D plot, both for the observations and the group centers.
- If there are chosen the first two discriminant axes, then a 2D graphic is to be plotted, see *Figure 2*.
- *Numerical example.* The following table contains data regarding 11 individuals (i1,..., i11), 2 predictor variables and 3 groups (a, b, c).

Discriminant Analysis

Result graphical representations.

- The observation table:

Id	*G*	*x1*	*x2*
i1	a	2	2
i2	a	3	1
i3	a	4	2
i4	b	6	4
i5	b	7	5
i6	c	8	1
i7	b	8	4
i8	b	9	5
i9	c	10	2
i10	c	11	1
i11	c	12	1

Discriminant Analysis

Result graphical representations.

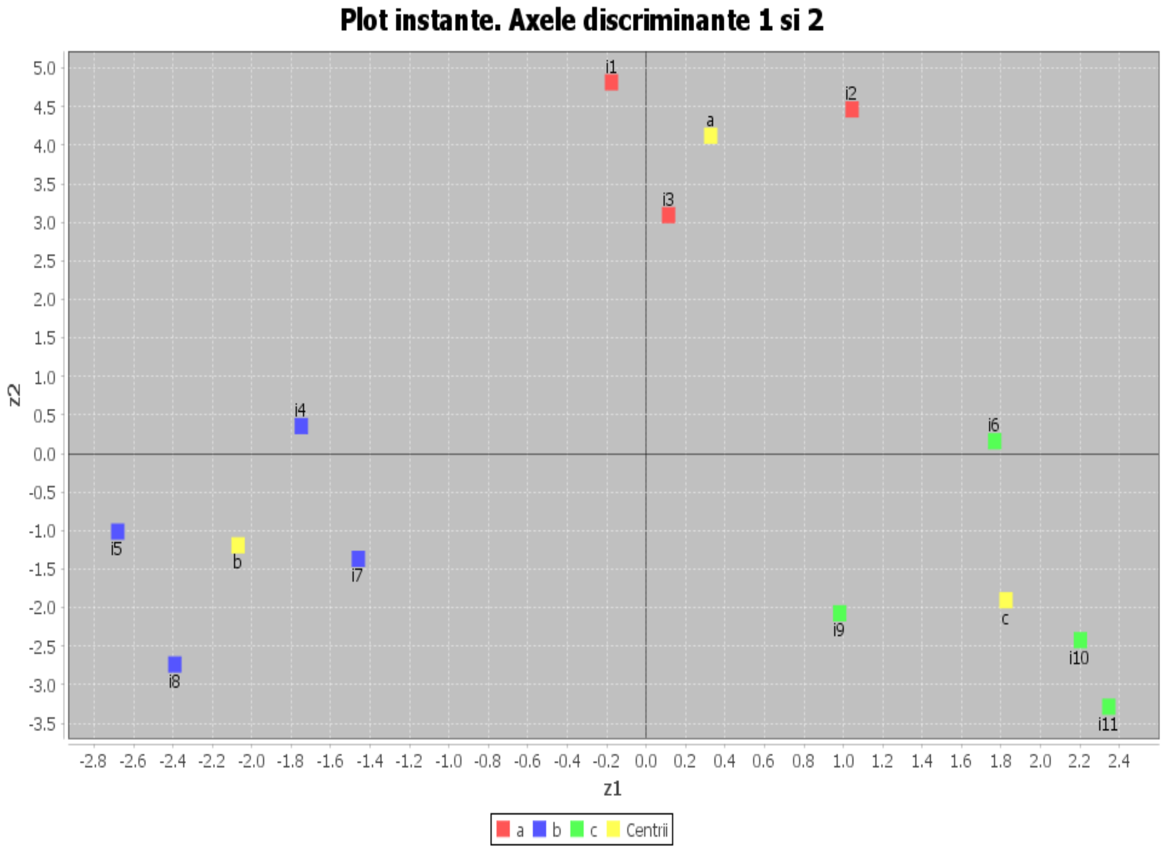


Figure 2*. Observations and group centers (n = 11, q=3)*

Discriminant Analysis

Result graphical representations.

- The discriminant variables are:

	z1	z2
i1	-0.17581	4.81766
i2	1.04474	4.465
i3	0.11348	3.09546
i4	-1.74904	0.35639
i5	-2.68031	-1.01315
i6	1.76796	0.15952
i7	-1.45976	-1.3658
i8	-2.39102	-2.73534
i9	0.98134	-2.07111
i10	2.20189	-2.42376
i11	2.34653	-3.28486

Discriminant Analysis

Result graphical representations.

- The discrimination power:

Nr.	λ_k	α_k
1	10.6361	0.91406
2	5.49765	0.8461

- Discrimination factors, and discrimination function coefficients:

	u_1	u_2
x1	0.14464	-0.8611
x2	-1.07591	-0.50844

Discriminant Analysis

Result graphical representations.

- The free terms are:

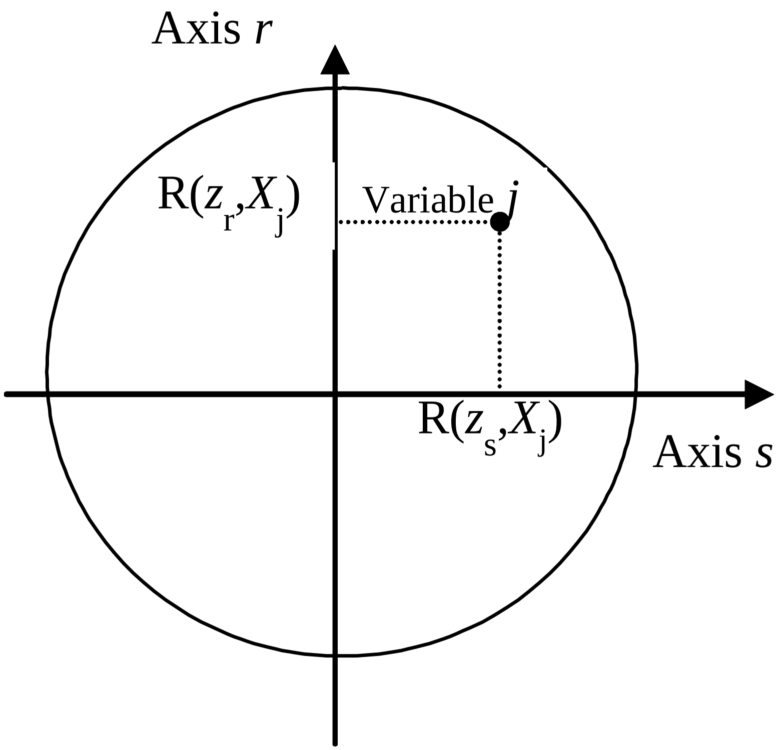
u_{01}	u_{02}
1.68672	7.55673

- *Variable representation.* The discriminant variables are not correlated to each other.

Discriminant Analysis

Result graphical representations.

- Using the correlation circle, the correlation between the predictor and discriminant variables can be represented.



Discriminant Analysis

Discriminant analysis as a particular case of canonical analysis.

- Since the first set of predictor variables contained in the observation table is represented by matrix *X*, and the second set of variables is given by the columns of the complete disjunctive matrix *Y*, built based on the discriminant variable.
- The canonic factor of order *k* from the first set is the eigenvector corresponding to order *k* eigenvalue of matrix $(V_{11})^{-1}V_{12}(V_{22})^{-1}V_{21}$, i.e.:

$$V_{11} = \frac{1}{n} X^t X \quad , \quad V_{22} = \frac{1}{n} Y^t Y \Rightarrow V_{22}^{-1} = nD_G^{-1}$$

$$V_{12} = \frac{1}{n} X^t Y = \frac{1}{n} G^t D_G \ , \ V_{21} = \frac{1}{n} Y^t X = \frac{1}{n} D_G G$$

Discriminant Analysis

Discriminant analysis as a particular case of canonical analysis.

- Resulting that:

$$V_{11}^{-1}V_{12}V_{22}^{-1}V_{21} = n(X^tX)^{-1}\frac{1}{n}G^tD_GnD_G^{-1}\frac{1}{n}D_GG = (X^tX)^{-1}G^tD_GG = T^{-1}B$$

- i.e. it is the matrix with the eigenvalues representing the discrimination power of discriminant variables, and having the eigenvectors the discrimination factors or the coefficients of the discrimination functions.
- Therefore, the canonical factors of the first set of variables (the set of predictor variables, *X*) are actually the discriminant factors.

Discriminant Analysis

Discriminant analysis as a particular case of canonical analysis.

- Canonical analysis allows for applying significance tests for each discriminant variable.
- The significance test for a canonical root shows in this context whether a discriminant variable is a good predictor.